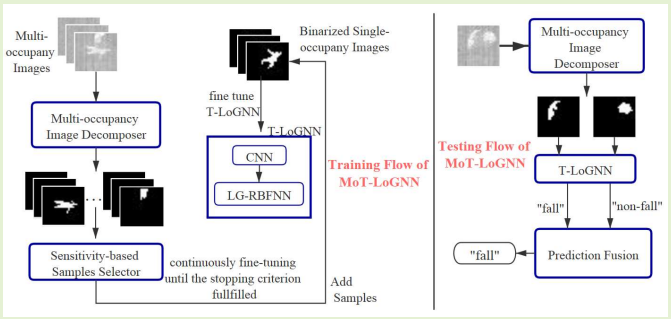
# Multi-occupancy Fall Detection using Non-Invasive Thermal Vision Sensor

# Multi-occupancy Fall Detection using Non-Invasive Thermal Vision Sensor

Cankun Zhong, Wing W. Y. Ng*, Shuai Zhang, Chris Nugent, Colin Shewell, Javier Medina-Quero
*Corresponding Author

***Abstract*—Falling is a common issue within the aging population. The immediate detection of a fall is key to guarantee early and immediate attention to avoid other potential immobility risks and reduction in recovery time. Video-based approaches for monitoring fall detection, although being highly accurate, are largely perceived as being intrusive if deployed within living environments. As an alternative, thermal vision-based methods can be deployed to offer a more acceptable level of privacy. To date, thermal vision-based fall detection methods have largely focused on single-occupancy scenarios, which are not fully representative of real living environments with multi-occupancy. This work proposes a non-invasive thermal vision-based approach of multi-occupancy fall detection (MoT-LoGNN) which discriminates between a fall or no-fall. The approach consists of four major components: i) a multi-occupancy decomposer, ii) a sensitivity-based sample selector, iii) the T-LoGNN for single-occupancy fall detection, and iv) a fine-tuning mechanism. The T-LoGNN consists of a robust neural network minimizing a Localized Generalization Error (L-GEM) and thermal image features extracted by a Convolutional Neural Network (CNN). Comparing to other methods, the MoT-LoGNN achieved the highest average accuracy of 98.39% within the context of a multi-occupancy fall detection experiment.**




***Index Terms*—Multi-occupancy Fall Detection, Thermal Vision Sensor, MoT-LoGNN, smart environments, Neural Networks**

## I. Introduction

As the population continues to grow on a global scale [1], increasing pressure is being placed on health and care services to meet the demands of increased numbers of persons requiring care provision. In such scenarios, those suffering from long term chronic conditions [2] or the aging population have the potential to benefit the most from the introduction of a new health-care delivery paradigm [3]. Besides, low-cost sensing solutions, whose wireless services coupled with rapid advances in data analysis, have provided the next generation of products to be deployed within living environments. These have the potential to improve the manner where remote health-care support can be provided and are slowly gaining increased acceptance by both users and health-care professionals [4].

From the multitude of health scenarios to consider, detecting falls within the living environment is a relevant challenge with a high impact in terms of both security and safety. Accidental falls can cause serious injury to at-risk individuals, especially for the aging [5]. Within this cohort, falls are the leading cause of hospitalization, injury-related deaths and loss of independence. However, it has been demonstrated that detecting and rapidly responding to falls can reduce the long-term risks associated with falls.

Although efforts have been directed towards supporting the detection and management of falls within living environments, a range of issues still exist. From a usability perspective, challenges are faced by the costs of the solution and the perceived issue of intrusiveness when video based cameras are used. From a technical perspective, challenges are faced by levels of accuracy levels and a desire to reduce the numbers of false positives given the implications that these have from a health-care provision perspective.

In addition, the studies of fall detection are mainly

This work was supported in part by 2020 R&D Program in Key Areas of Guangdong Province (No. 2020B010166002), the National Natural Science Foundation of China under Grant 61876066, and Guangdong Province Science and Technology Plan Project (Collaborative Innovation and Platform Environment Construction) 2019A050510006. In addition, this work was partially supported by the REMIND project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 734355.

Cankun Zhong and Wing W. Y. Ng (corresponding author) are with the Guangdong Provincial Key Laboratory of Computational Intelligence and Cyberspace Information, School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: curran.z@qq.com and wingng@ieee.org).

Shuai Zhang, Chris Nugent and Colin Shewell are with School of Computing, Ulster University, Northern Ireland, UK (e-mail: s.zhang@ulster.ac.uk, cd.nugent@ulster.ac.uk and cp.shewell @ulster.ac.uk).

Javier Medina Quero is with Department of Computer Science, University of Jaen, Jaen, Spain (e-mail: jmquero@ujaen.es).

focused on the single-occupancy scenario, because they think that in the multi-occupancy scenario, standing people can provide help for the fallen person. However, they may not be able to provide timely help for the fallen person. Alternatively, the machine actively detecting the occurrence of falls can contact the most professional medical staff in the vicinity at the first time. Besides, when there is an accident, bystanders may hardly be able to save themselves, not to mention that providing the help.

In this paper, we introduce a novel end to end solution for the remote management of falls within a multi-occupancy living environment. A low-cost sensing solution is presented, which has been developed based on the use of low-resolution thermal sensors. This configuration of sensors enables capturing activity in an unobtrusive manner and integrating data into a scalable sensor platform where an innovate approach for thermal image processing is deployed. The classification of fall or non-fall is computed in real-time using image decomposition and classification with a neural network (NN), which trained via minimization the Localized Generalization Error with features extracted by Convolutional Neural Networks (CNN). The developed approach has been deployed within 2 smart lab environments in the UK and Spain and has been evaluated by means of collection and analysis of labeled data sets.

The remainder of this paper is organized as follows. Section II provides a brief review of sensors which have been used for fall detection and approaches which address fall detection problems using thermal cameras. Section III describes the thermal camera used in this work and the proposed fall detection method. Section IV discusses the experimental settings in the smart labs and experimental results. Finally, conclusions are drawn and outlook for future work is presented in Section IV.

## II. Related Work

### A. Sensors for Fall Detection

A number of approaches have been implemented in an attempt to improve the process of detecting falls. From a sensing perspective, these have either been centered around exploiting wearable sensors and environmental sensing approaches [6].

From a wearable sensing perspective, efforts have been directed towards the processing of data gleaned through sensors such as accelerometers, gyroscopes, and barometers [7]. To further improve the detection accuracy, multi-sensor data can be utilized jointly as well [8]. More recently the sensing platforms within smart phones [9] or within smart shoes [10] have also been leveraged to detect falls. Although accuracy levels of detecting a fall have been reported to be in excess of 90% in some studies these solutions have a major disadvantage that they must be worn to offer their functionality. To a certain extent, this requirement can be viewed as being both an inconvenience and intrusive for the user and in some instances can be forgotten to be carried or in the worst case not used at all.

Approaches based on environmental sensors rely on the technology being deployed at mostly fixed locations. A wide variety of sensors have been proposed to describe the fall of a person, such as vibration detection sensors, video cameras, pressure sensors, and thermal sensors [11]. All of these approaches have their advantages and disadvantages.

Acoustic sensors can be used to detect a noise which is atypical of a fall event. These are, however, comprised in noisy environments where background noise interferes with the underlying sound of the fall [12]. Vibration sensors can be tuned to detect the measurement of a sudden impact which can be representative of a fall. They are, however, subject to false positives due to activities such as heavy walking in the environment. Video cameras provide what may be the only definitive solution to record what has happened in an environment [13]. Nevertheless, they suffer from the significant issue of perceived intrusiveness when they are deployed to monitor the daily activities of users in real homes. A potential alternative to video-based sensing is the use of thermal cameras. Due to the low resolution of thermal images the intrusiveness issue is overcome whilst at the same time having the ability to collect sufficient information from the heat of the human body to capture a fall event.

### B. Fall Detection Methods based on Thermal Cameras

Many methods have been applied in an attempt to improve the performance of automated fall detection based on thermal cameras. W. K. Wong, *et al* [14] utilized the width height ratio of the rectangle bounding the human as the feature and set up artificial rules to detect the falls. The x-, and y-axis histograms were utilized as input features for the SVM (Support Vector Machine) model to detect falls of patients [15]. The FallSense method was proposed in [17] which adopts fuzzy inference system based on the accleration, infrared, and ultrasonic snesors to detect falls. Experimental results show that the FallSense achieves overall 16% improvement in comparison with comparative methods on an average. P.Mazurek, *et al* [18] applied the traditional machine learning classifiers (i.e. support vector machine, artificial neural network, and naïve Bayes classifier) for fall detection using the kinematic features and mel-cepstrum-related features extracted from the thermal images. Experimental results show that the accuracy of the proposed method is more than 90% on two data sets.

A number of studies placed the thermal sensors on the ceiling as an alternative to a wall mounted solution in an effort to provide a broader view and to reduce occlusions. In [18], the thermal pixels of occupants were identified through a certain temperature range, and then the thermal pixel count of the occupant was used to detect the fall. Only focusing on the number of pixels made this method ignore the shape and edge of a detected person, thus affecting the fall recognition accuracy. In [19], the authors separated the foreground from the background based on temperature values. Manually set features based on temperature difference and temporal information were proposed and evaluated by several classifiers to detect falls. Experimental evaluation demonstrated that the system achieved real-time operation and over 94% fall recognition rate at room temperatures up to 24ºC.

As an alternative to feature extraction developed by statistical and/or knowledge based approaches, more recent approaches are focused on deep learning-based methods that automatically extract features from thermal videos for detecting falls. Studies [20]-[21] utilized the manually designed CNNs and Study [22] chose one of the most used Inception-v3 model which pre-trained on the ImageNet database to detect falls from the thermal images, where the classification accuracy were all higher than 80% in the single-occupancy scenario. The fall detection problem was treated as an anomaly detection problem in [23]-[24] where the deep learning framework Autoencoders were proposed to learn spatio-temporal features automatically from thermal videos. A fall was identified as an anomaly based on the reconstruction error.

The majority of the current studies have, however, focused on single-occupancy scenarios. The presentation of fall detection solutions in multi-occupancy scenarios has rarely been mentioned except for the recently conducted work [20]. Given the prevalence of such an event occurring, this area requires further attention. Fall detection solutions using thermal sensors considering single-occupancy scenarios may not only generate higher false alarm rates, however, may also fail to detect falls precisely in the context of a multi-occupancy scenario given that the feature distribution of the training (single-occupancy) data is quite different from the one of testing (multi-occupancy) data.

Thermal images collected from low-resolution thermal sensors are usually noisy and blurred and multi-occupancy thermal images are more difficult to recognize than single-occupancy thermal images. This is largely related to the additional people in the scene developing similar shapes as the person who has actually fallen when they are in close proximity of each other. In this research, we propose a multi-occupancy fall detection method MoT-LoGNN using thermal vision sensors. The major contributions of this work are summarized as follows:

1) To support multi-occupancy scenarios, we propose a robust fall detection method (MoT-LoGNN). Experimental results have demonstrated that the MoT-LoGNN yields the best performance in comparison to benchmarking techniques.
2) A stochastic sensitivity measure (SSM) is applied to both sample selection and validation for fine-tuning the T-LoGNN where the T-LoGNN consists of a robust neural network minimizing a Localized Generalization Error (L-GEM) and thermal image features extracted by a CNN. The SSM is a key component in the aforementioned L-GEM which measures the classifier sensitivities with respect to perturbations to input features. The minimization of the SSM and the L-GEM enhance the robustness of the MoT-LoGNN.
3) In this study, Radial Basis Function Neural Network trained via minimization of L-GEM is proposed as the classifier for fall detection to offer robust classification within the context of noisy and blurred thermal images.

## III. MoT-LoGNN

In this study, the proposed MoT-LoGNN consists of four components: the T-LoGNN, a fine-tuning mechanism, a multi-occupancy decomposer (MOD), and a sensitivity-based sample selector (SSS). The combination of the CNN and the LG-RBFNN is referred to as T-LoGNN, where the CNN is used as a feature extractor and a robust Radial Basis Function Neural Network trained via minimization of the Localized Generalization Error (LG-RBFNN) is used as a classifier. T-LoGNN is initially trained using the labeled single occupancy training set. The MOD decomposes multi-occupancy thermal images into one or more thermal single-occupancy sub-images, each only having a single person. Misclassified sub-images by T-LoGNN are selected by the SSS based on the sample sensitivity value to update the single-occupancy training set. The updated single-occupancy training set is then used for fine-tuning the T-LoGNN. It should be noted that if at least one sub-image decomposed from a multi-occupancy image is classified as fallen by the T-LoGNN, then this multi-occupancy thermal image will be classified as fallen as well.

The flow of process for the MoT-LoGNN training is presented in Fig. 1. Further details will be provided in the ensuing sub-sections.

For testing purposes, a multi-occupancy thermal image is input to the MoT-LoGNN. Firstly, the thermal image is decomposed into one or more (see further details in section II-C) single-occupancy thermal sub-images by the MOD. Following this, the T-LoGNN classifies all thermal sub-images. A multi-occupancy thermal image is classified as a fallen event in the instance that at least one of its thermal sub-images is classified as fallen.

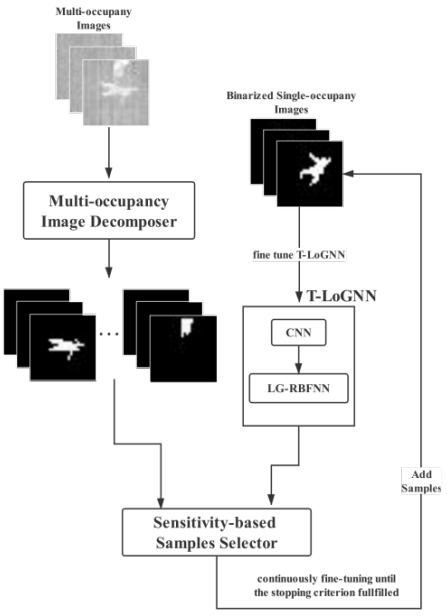



Fig. 1. Flow of processes for the MoT-LoGNN training

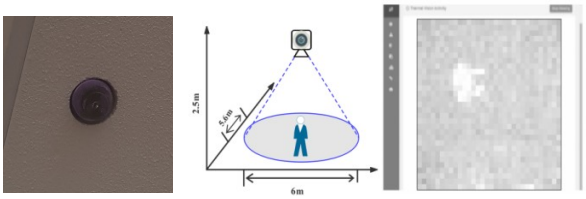
In the following sub-sections, the thermal vision sensor hardware is firstly introduced. Following this, four major components of the MoT-LoGNN: the MOD, the T-LoGNN, the fine-tuning mechanism, and the SSS are presented in Sections II-B, II-C-1, II-C-2, II-D, respectively. Finally, the time complexity analysis of the MoT-LoGNN in the testing phase is presented in II-E.

### A. Low Cost and Non-invasive Thermal Sensor

Between the range of thermal vision sensors [25], the high resolution [26] and low resolution devices are used in smart environments [27]. In the case of fall detection, a comparative of thermal sensor devices [28] has shown that non-invasive and low resolution thermal sensors have better performance and reduction of learning time.

Based on the encouraging results on fall detection from previous works, in this work we select the thermal sensor [20] Heimann HTPA 32x31, a suitable device with an operating temperature range of -20 to 85 ºC and powered by a 3.3 Volt supply. The thermal sensor generates a 32*31 matrix, where each value defines a heat point of temperature. The data are collected in real-time by means of a Ethernet crossover cable which is connected to the local area network. The middleware [29] collect and recover the data from the sensor in real time within a Web Service in JSON format.

As suggested in [20][28], the thermal sensor was affixed to the ceiling of the Smart Lab in the Ulster University to provide a zenith view of the space to be monitored. It was deployed at a height of 2.5 meters in a removable plaster ceiling, where the Ethernet connection and power supply keep hidden by the ceiling. It provides a viewable area of approximately 6 meters by 5.6 meters which makes it possible to monitor multiple individuals at the same time. The field of view of the sensor is 86° by 83°. A picture of the sensor deployed in the ceiling is provided in Figure 2(a) together with the operating range in Figure 2 (b) and an example of an actual recording in Figure 2 (c).




(a) (b) (c)

Fig. 2. (a) The sensor deployed in the ceiling. (b) The operating range of the sensor. (c) An example of a thermal image presented in a web interface.

Frames are sampled from the sensor through an $I^2C$ interface at a rate of 6 Hz and processed by a listener which communicates via wifi directly with endpoints on the SensorCentral platform [30]. Once captured by SensorCentral image processing techniques are invoked.

In this study, the aim is to identify whether a fall has occurred in the thermal images rather than consider what has happened to each individual in the thermal images. Therefore, the fallen or not labels were applied to the thermal frame images during the data collection, not a bounding-box label for each individual of each thermal image.

### B. Multi-occupancy Image Decomposer (MOD)

The MOD consists of three steps: 1) Image Binarization, 2) Contour Detection, and 3) Single-occupancy Thermal Sub-images Generation.

***1) Image Binarization:*** The image binarization process aims to distinguish the human heat points regarding the floor heat points using two pixel values: 0 and 255, respectively for floor and human shape. The determination of the binarization threshold influences both the decomposition result and the detection accuracy of the T-LoGNN. In this study, the binarization threshold is set to be 201 which is determined using a validation set. More details are presented in Section IV.

***2) Contours Detection***: The border-following algorithm in [31] is applied to find the contour of each person in the binarized thermal images. Owing to noise and blur in thermal images, very small contours are created which is likely to be mistaken by the high-temperature floor and are subsequently removed. For 28×28 thermal images that are cropped from the center of the original 32×31 images to contain the most relevant visual information, small contours whose areas are less than 4 pixels are removed since these areas would be too small to represent a human at the intended sensor deployment height.

***3) Single-occupancy Thermal Sub-images Generation*:** For each multi-occupancy thermal image, $k$ single-occupancy sub-images are generated if $k$ contours are found ($k$>1). If there is either zero or one contour found, the entire thermal image is treated as a thermal sub-image. For each contour, pixels located outside it are set to 0 while others are set to 255. Such that, a single-occupancy thermal sub-image is created and the detected person (i.e. contour) appears at the original location of the entire multi-occupancy thermal image. Besides, the fallen or not fallen class label of a thermal sub-image is inherited from this original multi-occupancy thermal image.

### C. T-LoGNN

The T-LoGNN consists of a robust LG-RBFNN and thermal image features extracted by a CNN. The LG-RBFNN and the fine-tuning mechanism [32] for T-LoGNN are introduced in II-B.1 and II-B.2, respectively.

The CNN is adopted in the T-LoGNN for feature extraction from both binarized thermal sub-images and single occupancy images. One of the major contributions of this work is the use of the LG-RBFNN as the classifier with features extracted from the CNN. In contrast to the Softmax classifier, the LG-RBFNN is expected to yield higher generalization capability to future unseen samples since it is trained via minimizing the generalization error estimated by the L-GEM. Furthermore, RBFNN is used here because it is a nonlinear classifier [33] with fast convergence.

A class balanced weighting trick is utilized in the training of the CNN when the model performance is hindered by the class imbalance problem. The class weight of each class is the reciprocal of its number of samples multiplied by a constant.

***1) LG-RBFNN:*** The purpose of RBFNN training is to find a network structure and connection weights to minimize the generalization error, such that it will be more robust to the effects of noise and blurred areas in thermal images. Once the number of hidden neurons are determined, the centers and widths of hidden neurons can be obtained by k-means clustering. After fixing both the centers and widths, connection weights can be calculated by a pseudo-inverse technique. Therefore, the objective of RBFNN training can be simplified to the finding of the optimal number of hidden neurons which minimizes the generalization error. We cannot, however, directly estimate the generalization error. In this study, the L-GEM model proposed in [34] is used to find the upper bound of generalization errors of the RBFNN.

The L-GEM defines the generalization error of unseen samples located near training samples only, i.e. $Q$-union ($S_Q$). The concept is that generalization errors of unseen samples with a large difference from the training samples are expected to be large because we have no knowledge of such unseen samples. Therefore, estimating generalization errors for unseen samples far away from the training samples may be counterproductive and misleading [35].

$S_Q$ is the union of $Q$-neighborhoods of all training samples and a $Q$-neighborhood ($S_Q(x_f)$) of a training sample $x_f$ (feature vector extracted by CNN) is a local input space which includes all unseen samples located near $x_f$. The $S_Q(x_f)$ is defined as follows:

$$S_Q(x_f) = \{x | x = x_f + \Delta x, |\Delta x_i| \le Q, i = 1,2,...,d\} \quad (1)$$

where $\Delta x = (\Delta x_1,...,\Delta x_d)'$, $\Delta x_i$, and $i$ denote the stochastic perturbation, the stochastic perturbation on the $i^{\text{th}}$ input feature, and the number of input features, respectively. According to [28], for a given $Q$ value, the upper bound of the LGEM ($R^*_{SM}(Q)$) is estimated by using the Hoeffiding's inequality with a probability of $1-\eta$. The definition of $R^*_{SM}(Q)$ is as follows:

$$R^*_{SM}(Q) = \left(\sqrt{R_{emp}} + \sqrt{E_{SQ}((\Delta y)^2)} + A\right)^2 + \varepsilon \quad (2)$$

where $\varepsilon = B\sqrt{\ln\eta/(-2N)}$, $N$, $A$, $B$, $R_{emp}$, and $\sqrt{E_{SQ}((\Delta y)^2)}$ denote the number of training samples, the difference between the maximum and the minimum value of outputs, the minimum value of training mean square error, the training mean square error, and the SSM of output differences, respectively.

Note that both $\varepsilon$ and $A$ are constants for a given training dataset. Let $g(\cdot)$ be the classifier and the definition of the SSM is the expectation of squares of classifier output perturbations ($\Delta y = g(x_f + \Delta x) - g(x_f)$) between training samples and unseen samples in $S_Q$:

$$E_{SQ}((\Delta y)^2) = \frac{1}{N}\sum_{f=1}^{N} E[(g(x_f + \Delta x) - g(x_f))^2] \quad (3)$$

In general, we do not have any prior knowledge about the distribution of unseen samples in $S_Q$, thus we adopt a quasi-Monte Carlo (QMC) based method as in [35] to estimate the SSM value as follows:

$$\begin{aligned} E_{SQ}((\Delta y)^2) &= \frac{1}{N}\sum_{f=1}^{N} SSM(x_f, g) \\ &= \frac{1}{N}\sum_{f=1}^{N}\left(\frac{1}{H}\sum_{h=1}^{H}(g(x_f + \Delta x_h) - g(x_f))^2\right) \end{aligned} \quad (4)$$

where $SSM(x_f,g)$, $H$, $\Delta x_h$ denotes the sensitivity measure of the classifier $g$ for training sample $x_f$, the number of Halton points, a Halton point where each coordinate ranges from $[-Q,Q]$, respectively.

By fixing $Q$, the optimal RBFNN is found by searching for the optimal number of hidden neurons which yields the minimum generalization error. Procedures for finding the optimal RBFNN are presented in Algorithm 1.

**Algorithm 1** Finding optimal RBFNN

**Input:** $x$, $Q$
**Output:** optimal RBFNN
1: **for** $M \leftarrow$ number of classes to $N-1$ do
2: Train an RBFNN with $M$ hidden neurons using $x$.
3: Compute the $R^*_{SM}(Q)$ for the current RBFNN.
4: **end for**
5: **return** RBFNN yielding the minimum $R^*_{SM}(Q)$

Given the fact that the temperature of the floor in the entire monitoring area may not be uniform such that white (high temperature) dots may be added to the background as noise in cases. These perturbations (noises) introduced to the region of interest (i.e. fallen person) may increase the difficulty of fall detection since it will cause the activation values of neurons in CNN to become larger or smaller and thus affect the prediction result. In T-LoGNN, The SSM measures fluctuations of classifier outputs with respect to input (feature vector) perturbations. Therefore, the RBFNN trained via minimizing both SSM value and training error (T-LoGNN) is more robust to the noise present within the thermal images.

***2) Fine-tuning Mechanism of T-LoGNN:*** The T-LoGNN is pre-trained using a set of binarized single-occupancy thermal images for classifying whether or not the person in the image has fallen. Prior to training the MoT-LoGNN, a set of multi-occupancy thermal images for training is divided into 80% for constituting a multi-occupancy training set and the remaining 20% for forming a multi-occupancy validation set. These thermal images for training are subsequently decomposed into single-occupancy thermal sub-images by the MOD. Training begins with the T-LoGNN classifying each thermal sub-image generated from the training set. Those misclassified or possibly misclassified thermal sub-images are selected by the SSS and added to the single-occupancy training set for fine-tuning of the T-LoGNN. The process iterates until the limit of iterations ($n$) is reached. In our experiments, it is found that when $n$ is less than or equal to 5, the classification accuracy of the thermal images on the training set can be guaranteed to be 100%. Therefore, in this experiment, $n$ is set to 5.

To prevent overfitting, the most robust T-LoGNN among $n$ iterations is selected as the final T-LoGNN to be used in the MoT-LoGNN. A validation robustness measure (VRM) is proposed to measure the robustness of the T-LoGNN as follows:

$$\frac{1}{V}\sum_{v=1}^{V}(y_v - F(X_v)) + \frac{1}{W}\sum_{w=1}^{W} SSM(X_w, f) \tag{5}$$

where $F(\cdot)$, $f(\cdot)$, $V$ , $W$, $X_v$, $y_v$, $X_w$ denote the MoT-LoGNN, the T-LoGNN, the number of samples in the multi-occupancy validation set, the number of single-occupancy sub-images generated from the multi-occupancy validation set, a multi-occupancy image from the validation set, the label of $X_v$, a single-occupancy sub-image generated from the multi-occupancy validation set, respectively. The best T-LoGNN yielding the minimum VRM needs to yield both low classification error on the validation set and low sensitivity to small perturbations to thermal sub-images. This follows the idea of the minimization of the L-GEM in [33]-[34] and aligns well with the multi-occupancy falling classification problem in this work.

### D. Sensitivity-based Samples Selector (SSS)

An SSS approach is proposed to select useful samples for fine-tuning the T-LoGNN. The main difficulty is to identify misclassified single-occupancy thermal sub-images since the class label is assigned to the entire multi-occupancy thermal image only and we do not have the real label of the sub-images.

The three following possible cases are considered by SSS to select useful sub-images for fine-tuning T-LoGNN, which are presented in Fig. 3.

Both Cases 1 and 2 identify misclassified single-occupancy scenarios whilst Case 3 provides additional samples for people who have not fallen.

In Case 2, multiple sub-images are classified as non-fallen, but at least one of the sub-images should be classified as fallen. In Case 3, multiple sub-images are classified as fallen, but the classification of multiple fallen sub-images from the same multi-occupancy thermal image has a high chance to be mistaken because both fallen events existing at the same time is quite rare. However, we do not have a bounding-box label for each participant in the dataset, therefore it is not possible to know which sub-image is misclassified.

1) For non-fallen (i.e. all individuals in a multi-occupancy thermal image are standing or no one in a thermal image) multi-occupancy thermal images misclassified as fallen. The sub-images which have been classified as fallen are incorrect and are subsequently selected;
2) For fallen thermal images misclassified as non-fallen, sub-images yielding the largest SSM value among sub-images decomposed from the same multi-occupancy thermal image are selected;
3) For correctly classified fallen multi-occupancy thermal images with more than one sub-image being classified as fallen, the thermal sub-image yielding the largest stochastic sensitivity measure value among the sub-images being classified as fallen and decomposed from the same thermal image are selected.

In the L-GEM framework, a sample yielding a large SSM value is informative to Neural Network training [36] because it has a higher chance of being misclassified by the Neural Network. Therefore, for Cases 2 and 3, the sub-image yielding the largest SSM ($SSM(x_f, g)$ in Equation (4)) is selected. Selected samples are labeled with the opposite labels with respect to their corresponding classification results from the T-LoGNN. Then, these samples are added to the ensuing round of fine-tuning T-LoGNN.

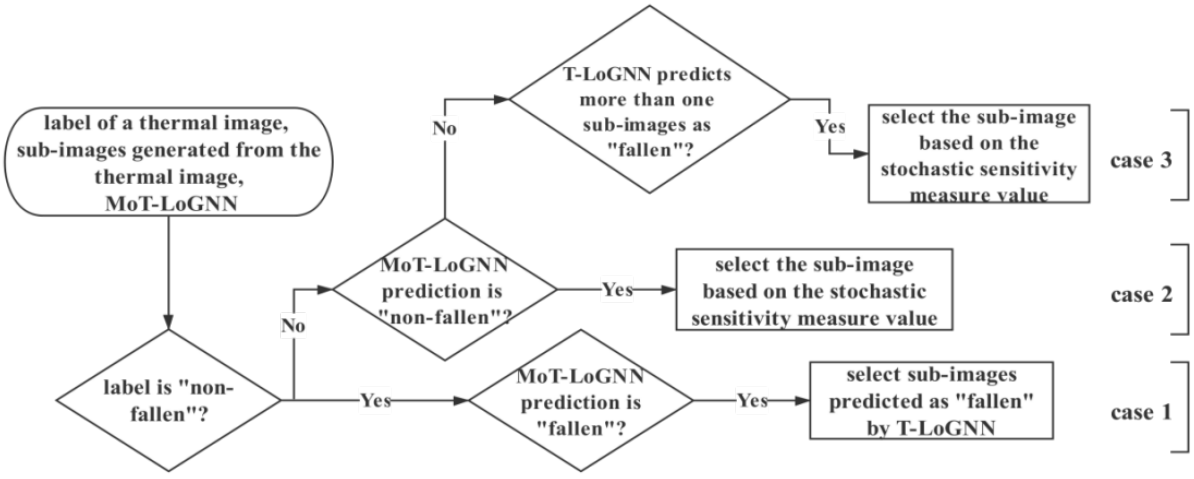



Fig. 3. Informative sub-images selected by SSS.

### E. Time Complexity Analysis of MoT-LoGNN in testing phase

In the testing phase, only two components (MOD and T-LoGNN) of the MoT-LoGNN are involved. Their time complexity is given as follows.

For the MOD consisting of three steps, time complexities of all steps are all $O(ij)$, where $i$ and $j$ denote the width and the height of a thermal image, respectively. Therefore, the total time complexity of MOD is $T_{MOD} = O(ij)$.

For the T-LoGNN, the total time complexity of all convolutional layers in the CNN is $T_{CNN} = O(\sum_{l=1}^{L} c_{l-1} c_l s_l^2 m_l^2)$ Here $L$, $l$, $s_l$, $m_l$, $c_{l-1}$ and $c_l$ denote the number of convolutional layers, the index of a convolutional layer, the size of the convolutional kernel, the size of a feature map of single channel in the $l^{th}$ layer, the number of input channels and the number of output channels of the $l^{th}$ layer, respectively. As to the pooling layers and fully connected layers of CNN, they often take 5% to10% computational time [37]. Furthermore, the time complexity of the LG-RBFNN is $T_{LG\text{-}RBFNN} = O(dMo)$, where $d$, $M$ and $o$ denote the dimension of the feature vector extracted by CNN, the number of hidden neurons, and the number of outputs, respectively.

Therefore, the time complexity of the MoT-LoGNN for classifying a thermal image is $T_{MOD}+T_{CNN}+T_{LG\text{-}RBFNN}$.

## IV. Experiments And Results

The Experimental setup is given in Section III-A. The MoT-LoGNN compared with other thermal fall detection studies is conducted in Section III-B. Some discussions of the details of MoT-LoGNN are presented in Section III-C.

### A. Experimental Setup

***1) Collected Data:*** The data for this experiment have been collected from the Smart Lab in the Ulster University. The thermal vision sensor was placed on the ceiling of the room to collect a zenithal view of the occupants, which provided a clear view of falls and also reduced the potential of occlusion in instances of multi-occupancy, compared with if a camera was installed in the vertical plane. Three participants aged between 25 and 35 (including one woman and two men) with heights 1.68, 1.72, and 1.83 meters, respectively assisted in collecting the data. When collecting data, each participant walked around the area within the field of view of the thermal vision sensor for walking scenarios and laid on the floor with changing the orientation, rotating, moving, bending the joints to simulate different scenarios of people who have fallen. An external observer collected frames from the thermal sensor in real-time and labeled each frame (fallen or non-fallen) manually during the development of scenes.

Each collected thermal image is pre-processed as a 28×28 matrix where each pixel defines a heat point with a value between 0 and 255. There are two sets of experimental data: single-occupancy and multi-occupancy. The single-occupancy dataset consists of three categories: (i) empty room, (ii) standing/walking alone, and (iii) fallen alone. For the multi-occupancy dataset two further scenarios were added: (iv) 2-3 people standing/walking, (v) one person fallen and one standing/walking. In the experiments, thermal images for both datasets in categories (iii) and (v) are labeled to be fallen while the remainder is labeled to be non-fallen. In experiments, for the collected single-occupancy dataset, numbers of samples being labeled as fallen and non-fallen are 186 and 159, respectively. For the collected multi-occupancy dataset, numbers of samples being labeled as fallen and non-fallen are 528 and 431, respectively. Collecting data from different inhabitants and cases of fallen take a great deal of efforts, so the data augmentation techniques (i.e. translation, flipping, and rotation) are utilized in this study to enlarge the number of samples by 100 times, which also helps to overcome the overfitting problem to some extent.

***2) Implementation Details:*** In the experiments, a 3-layer carefully designed CNN was employed for feature extraction from the thermal images to provide 1024-dimensional feature vectors. This CNN feature extractor is denoted by *cnn* and its architecture is presented in Fig. 4. It is optimized using the adaptive moment estimation method with learning rate = 0.001, $\beta_1$ = 0.9, $\beta_2$ = 0.999, and mini-batch = 32. Both Softmax (*softmax*) and RBFNN trained without L-GEM (*rbfnn*) are used in experiments for validating the robustness of the LG-RBFNN. The number of hidden neurons of the *rbfnn* was determined by minimizing the training classification error instead of the localized generalization error in LG-RBFNN. The learned features from the CNN were in the range [0, 3.85]. Values of $Q$ = 1 and $H$ = 100 were selected by a trial-and-error method and were determined by the trade-off between accurate estimation and computational time, respectively.

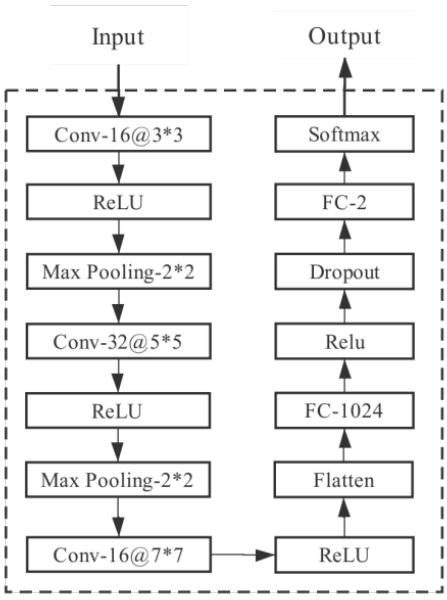



Fig. 4. The architecture of the CNN. Conv-k@n*n means the number of n*n convolution is k. All the kernel size of Max Pooling layer is 2*2. FC-1024 refers to a fully connected layer with 1024 neurons.

For both multi-occupancy and single-occupancy datasets, ten independent runs were performed for all experiments with 80% of thermal images being randomly selected for training and the remaining 20% used for testing.

***3) Evaluation metrics:*** The performance of different models is validated on the single-occupancy dataset and the multi-occupancy dataset. The mean and the standard deviation values of different metrics for different models are calculated and the symbol "*" denotes a statistically significant difference between the MoT-LoGNN and the corresponding method by Student's t-test with 95% confidence. We regard fallen as the positive class and non-fallen as the negative class. Three performance metrics are used. The Accuracy measures the overall performance of different models. The False Rejection Rate (FRR) and the False Acceptance Rate (FAR) measure the missing report rate of fallen and the false alarm rate, respectively. These performance metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{6}$$

$$\text{FRR} = \frac{FN}{FN + TP} \tag{7}$$

$$\text{FAR} = \frac{FP}{FP + TN} \tag{8}$$

$$\text{F1} = \frac{2*\left(\text{Precision}*\text{Recall}\right)}{\text{Precision} + \text{Recall}} \tag{9}$$

$$\text{Gmean} = \sqrt{\text{Recall} * \text{Specificity}} \tag{10}$$

where *TP, TN, FP,* and *FN* denote the True Positive, the True Negative, the False Positive, and the False Negative, respectively. Precision=*TP*/(*TP*+*FP*), Recall=*TP*/(*TP*+*FN*), and Specificity=*TN*/(*TN*+*FP*).

## B. Comparison Test with Other Fall Detection Studies using Thermal Sensor

In this section, the proposed MoT-LoGNN is compared with two artificial feature extraction based methods (the RectFall [14] using the width height ratio of the rectangle bounding the human as the feature and the HistFall [15] using the histogram of x-axis and y-axis as the feature) and two deep learning based methods (the manually designed CNN [21] and the popular Inception-v3 [22] model which pre-trained on the ImageNet database).

***1) Performance on single-occupancy dataset:*** It can be seen from table I that in the single-occupancy scenario, compared with other thermal sensor based fall detection methods, the proposed MoT-LoGNN achieves the best performance in terms of all metrics except the FRR. As tothe FRR, the MoT-LoGNN reaches the second lowest, while the RectFall with the lowest FRR has a high FAR of 40.62%, which means that the system is easy to make error alarm, resulting in waste of resources. Therefore, the MoT-LoGNN is obviously superior to other comparison methods in the single-occupancy scenario.

TABLE I
MEAN (± STDEV) OF PERFORMANCE METRICS OF DIFFERENT METHODS ON THE SINGLE-OCCUPANCY DATA

| Model | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean(%) |
|---|---|---|---|---|---|
| RectFall[14] | 86.43 (±3.69) | 40.62 (±5.63) | **0.00 (±0.00)** | 74.35 (±6.53) | 76.97 (±3.81) |
| HistFall[15] | 92.95 (±0.22) | 6.27 (±1.46) | 7.29 (±0.35) | 89.45 (±1.76) | 93.22 (±0.56) |
| CNN[21] | 91.80 (±2.95) | 13.15 (±4.31) | 4.37 (±1.98) | 87.53 (±3.09) | 90.15 (±3.44) |
| Inception-v3 [22] | 95.63 (±1.47) | 5.45 (±2.23) | 2.5 (±1.88) | 94.26 (±2.15) | 96.01 (±2.74) |
| MoT-LoGNN | **97.31 (±1.33)** | **0.91 (±0.48)** | 3.50 (±2.00) | **95.63 (±2.75)** | **97.93 (±0.79)** |

***2) Performance on multi-occupancy dataset:*** Since the RectFall and the HistFall methods are proposed for single-occupancy scenarios, they cannot be directly applied to the fall detection problems in multi-occupancy scenarios. In this experiment, the proposed MOD proposed in this paper is combined with these two methods, so that these two fall detection methods can be applied to multi-occupancy scenarios. As shown in Table II, in multi-occupancy scenarios, the performance of deep learning based fall detection methods are obviously better than the RectFall and the HistFall. Moreover, the FARs of the RectFall and HistFall methods are significantly higher than other methods. This is mainly due to the characteristics of the MOD framework. That is, as long as a thermal sub-image is classified as fall, it is considered that there is a fall. Therefore, the RectFall and the HistFall with high FAR in the single-player scenario have their disadvantages further expanded under the multi-occupancy scenario and the MOD. Overall, MoT-LoGNN has the highest average classification accuracy in multi-occupancy scenarios, the lowest average FAR, and the second lowest average FRR, but its FAR is dozens of times lower than the RectFall which has the lowest FRR.

TABLE II
MEAN (± STDEV) OF PERFORMANCE METRICS OF DIFFERENT METHODS ON THE MULTI-OCCUPANCY DATA

| Model | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean(%) |
|---|---|---|---|---|---|
| RectFall[14] | 64.46 (±1.36) | 79.38 (±1.47) | **0.12 (±0.09)** | 34.13 (±2.04) | 45.36 (±1.62) |
| HistFall[15] | 75.72 (±2.32) | 48.60 (±7.00) | 4.56 (±1.91) | 65.10 (±5.64) | 69.84 (±4.50) |
| CNN[21] | 85.46 (±0.77) | 13.15 (±4.31) | 15.42 (±4.53) | 84.19 (±0.99) | 85.60 (±4.48) |
| Inception-v3 [22] | 92.12 (±0.69) | 8.35 (±3.91) | 7.33 (±3.65) | 91.65 (±1.12) | 92.16 (±3.88) |
| MoT-LoGNN | **95.89 (±0.50)** | **4.12 (±1.32)** | 3.89 (±1.07) | **95.42 (±0.55)** | **95.92 (±0.68)** |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2020.3032728, IEEE Sensors Journal



### C. Discussions of MoT-LoGNN

***1) Sensitivity Analysis of MoT-LoGNN with Different Amount of Training Samples:*** The MoT-LoGNN extracts features using the CNN model. It is well-known that the amount of training data influences the performance of CNN. In this section, 10% and 50% of training samples are randomly selected to train the MoT-LoGNN to show the sensitivity of MoT-LoGNN with respect to the number of training samples. Tables III and IV show performances of the MoT-LoGNN with different amount of training data under the single-occupancy scenario and the multi-occupancy scenario, respectively. When the number of training samples increases, the performance of the MoT-LoGNN increases in both single-occpancy and multi-occupancy scenarios. The performance of models on the multi-occupancy dataset is of significant interest because it covers more possible scenarios than the single-occupancy dataset and is more representative to real living environments with multi-occupancy. As seen from Table IV, the performance of the MoT-LoGNN using all (100%) training data is a little better than the one using 50% number of training data on the multi-occupancy dataset in terms of the comprehensive metrics (no more than 0.6% improvement in terms of the Accuracy, F1, and Gmean metrics). It shows that when the number of training samples is large enough, the performance of the MoT-LoGNN is not sensitive to the amount of the training samples.

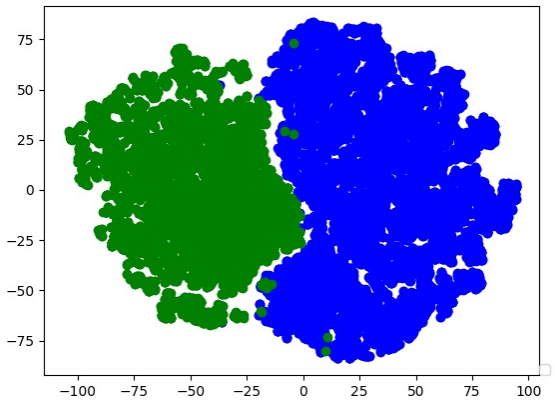
TABLE III
SENSITIVITY ANALYSIS OF MOT-LOGNN TRAINING WITH DIFFERENT AMOUNT OF TRAINING DATA ON SINGLE-OCCUPANCY SCENARIO

| Amount of Training Data | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean (%) |
|---|---|---|---|---|---|
| 10% | 91.72 (±2.44) | 11.60 (±6.20) | 6.00 (±2.41) | 87.62 (±2.43) | 91.09 (±3.27) |
| 50% | 96.40 (±1.51) | 1.06 (±1.10) | 4.63 (±1.92) | 94.26 (±3.21) | 97.14 (±1.02) |
| 100% | **97.31 (±1.33)** | **0.91 (±0.48)** | **3.50 (±2.00)** | **95.63 (±2.75)** | **97.93 (±0.79)** |

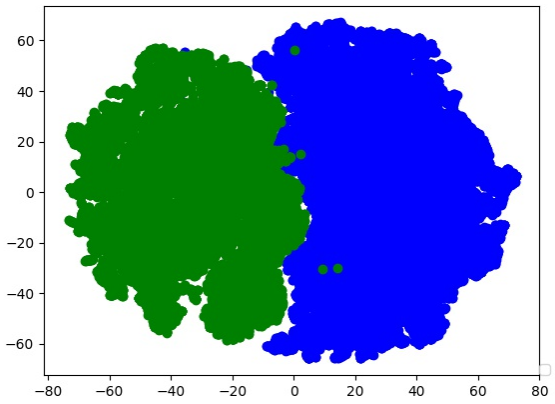
TABLE IV
SENSITIVITY ANALYSIS OF MOT-LOGNN TRAINING WITH DIFFERENT AMOUNT OF TRAINING DATA ON MULTI-OCCUPANCY SCENARIO

| Amount of Training Data | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean (%) |
|---|---|---|---|---|---|
| 10% | 83.83 (±7.90) | 24.04 (±20.53) | 8.71 (±3.75) | 79.49 (±12.01) | 82.16 (±10.43) |
| 50% | 95.44 (±0.29) | 4.26 (±1.19) | 4.68 (±1.51) | 94.92 (±3.21) | 95.52 (±1.96) |
| 100% | **95.89 (±0.50)** | **4.12 (±1.32)** | **3.89 (±1.07)** | **95.42 (±0.55)** | **95.92 (±0.68)** |

***2) Non-linearity of the Thermal Image Data:*** In this study, the CNN is utilized to extract features of thermal images and the LG-RBFNN serves as a non-linear classifier. To show the non-linearity of thermal image data, the data is transformed into a new feature space by a CNN and then it is reduced to a two-dimensional space by using the t-Distributed Stochastic Neighbor Embedding (tSNE), as shown in Fig. 5. Fig. 5 shows that the thermal image data in the feature space constructed by the CNN is non-linear separable in both single-occupancy and multi-occupancy scenarios. It is difficult to find a linear decision plane that separate fallen and non-fallen completely. The geometry of the optimal decision function in this two-dimensional feature space should be a curve instead of a straight line.




(a) Single-occupancy




(b) Multi-occupancy

Fig. 5. The data distribution of thermal images. The green points denote the non-fallen class and the blue points denote the fallen class.

Due to this non-linearity characteristic of the thermal sensor data, a non-linear classifier is preferred. Tables V and VI summarize the performance of different classifiers under the single-occupancy and multi-occupancy scenarios, respectively. Where the [*cnn+softmax*], [*cnn+dt*], and [*cnn+svm*] denote the models using the softmax, decision tree, and support vector machine as a classifier with the features extracted by the CNN, respectively. In both cases, the T-LoGNN yields the best performance which shows the superiority of the LG-RBFNN for dealing with the thermal sensor fall detection problem using the CNN features.

TABLE V
COMPARISON OF DIFFERENT CLASSIFIERS USING CNN FEATURES ON SINGLE-OCCUPANCY DATA

| Model | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean(%) |
|---|---|---|---|---|---|
| [*cnn+softmax*] | 92.80 (±0.76) | 10.72 (±5.90) | 4.71 (±2.61) | 88.91 (±0.65) | 92.14 (±1.97) |
| [*cnn+dt*] | 94.02 (±1.96) | 11.63 (±4.86) | 2.73 (±0.69) | 90.88 (±1.91) | 92.68 (±2.54) |
| [*cnn+svm*] | 93.71 (±0.89) | 12.04 (±5.15) | 2.82 (±1.66) | 90.12 (±0.08) | 92.39 (±1.88) |
| T-LoGNN | **95.10 (±1.88)** | **9.92 (±5.22)** | **1.91 (±0.41)** | **92.58 (±1.88)** | **93.95 (±2.57)** |

TABLE VI
COMPARISON OF DIFFERENT CLASSIFIERS USING CNN FEATURES ON MULTI-OCCUPANCY DATA

| Model | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean(%) |
|---|---|---|---|---|---|
| [*cnn+softmax*] | 88.77 (±1.49) | 13.31 (±2.40) | 9.42 (±1.65) | 87.37 (±1.40) | 88.60 (±1.48) |
| [*cnn+dt*] | 90.16 (±0.83) | 11.85 (±2.29) | 8.09 (±0.98) | 88.90 (±0.68) | 90.00 (±0.90) |
| [*cnn+svm*] | 89.98 (±1.42) | 11.59 (±1.39) | 8.59 (±2.82) | 88.74 (±1.65) | 89.88 (±1.33) |
| T-LoGNN | **90.81 (±0.43)** | **11.55 (±2.21)** | **7.07 (±2.83)** | **89.57 (±0.62)** | **90.63 (±0.26)** |

***3) Determination of the Binarization Threshold:*** As one of the components of the MoT-LoGNN, MOD is responsible for decomposing the multi-occupancy into single-occupancy sub-images, where the determination of the binarization threshold will influence the effect of the MOD directly. Intuitively, the optimal binarization threshold should separate the person and background clearly, which is helpful for the model to distinguish the fallen and non-fallen samples. In this study, a validation set (20% data randomly selected from the single-occupancy training set) is utilized to determine the binarization threshold. Furthermore, the T-LoGNN model is used as the fall detection model. The optimal threshold is determined once this threshold helps the T-LoGNN model reach the best performance in terms of the Gmean metric.

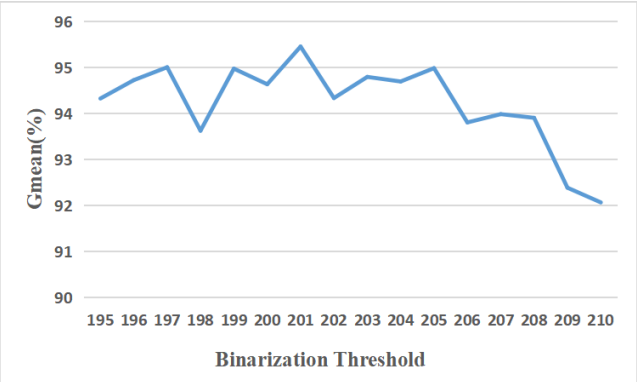



Fig. 6. The performance of the T-LoGNN using different MODs with different binarization thresholds.

In experiments, candidates of the binarization threshold ranges from 195 to 210 according to the prior experience. Fig. 6 shows that 201 is the optimal binarization threshold because it yields the highest Gmean. Therefore, the binarization threshold for the MOD is set to be 201 in this study.

***4) Ablation Study:*** Major contributions (i.e. the robust T-LoGNN for single-occupancy fall detection, the MOD for simplifying the multi-occupancy fallen detection into the simpler single-occupancy scenario, the SSS for enhancing the performance of the proposed T-LoGNN) of this work have been evaluated by experiments. Their results are shown in Tables VII and VIII.

TABLE VII
MEAN (± STDEV) OF PERFORMANCE METRICS OF DIFFERENT METHODS ON THE SINGLE-OCCUPANCY DATA

| Model | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean(%) |
|---|---|---|---|---|---|
| T-LoGNN | 95.10 (±1.88) | 9.92 (±5.22) | 1.91 (±0.41) | 92.58 (±1.88) | 93.95 (±2.57) |
| MOD-LoGNN | 97.09 (±0.28) | 1.70 (±2.00) | 3.66 (±2.08) | 95.24 (±2.81) | 97.50 (±0.63) |
| MoT-LoGNN | **97.31 (±1.33)** | **0.91 (±0.48)** | **3.50 (±2.00)** | **95.63 (±2.75)** | **97.93 (±0.79)** |

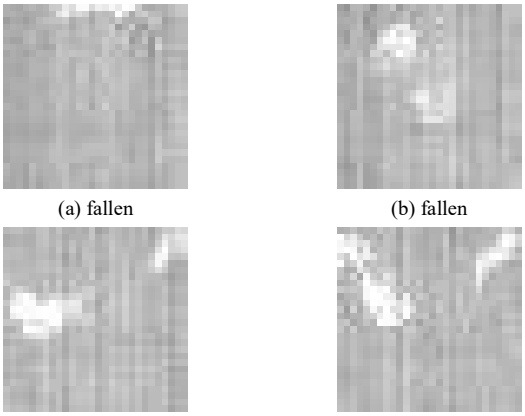
TABLE VIII
MEAN (± STDEV) OF PERFORMANCE METRICS OF DIFFERENT METHODS ON THE MULTI-OCCUPANCY DATA

| Model | Accuracy (%) | FAR (%) | FRR (%) | F1(%) | Gmean(%) |
|---|---|---|---|---|---|
| T-LoGNN | 90.81 (±0.43) | 11.55 (±2.21) | 7.07 (±2.83) | 89.57 (±0.62) | 90.63 (±0.26) |
| MOD-LoGNN | 92.57 (±1.10) | 12.40 (±1.43) | **0.85 (±0.62)** | 91.08 (±1.06) | 91.57 (±1.06) |
| MoT-LoGNN | **95.89 (±0.50)** | **4.12 (±1.32)** | 3.89 (±1.07) | **95.42 (±0.55)** | **95.92 (±0.68)** |

From Tables VII and VIII, after combining the MOD with the T-LoGNN, the overall performance (i.e. Accuracy, F1, and Gmean) is improved compared with the T-LoGNN only. The main reason is that the MOD performs an additional preprocessing of image binarization which can be regarded as a data denoising process. However, the FRR of MOD-LoGNN is higher than the one of the T-LoGNN on the single-occupancy data, which is mainly caused by the static binarization setting in this study. Therefore, the edge details of the fallen person with body curling may be filtered after the image binarization. In contrast, the MOD-LoGNN yields a higher FAR than the T-LoGNN on the multi-occupancy dataset. This is one of the drawbacks of the proposed method where a thermal image is claimed to be fallen as long as either one of its sub-image being classified as fallen.

The proposed MoT-LoGNN yields the highest Accuracy, F1, and Gmean on both single-occupancy and multi-occupancy dataset, which confirms the effectiveness of using SSS to select informative samples for fine-tuning the T-LoGNN. However, after using the fine-tuning mechanism (from MOD-LoGNN to MoT-LoGNN), the MoT-LoGNN yields a higher FRR. It may be caused by too many non-fallen samples are selected by the SSS to fine-tune the T-LoGNN and the class balanced weight technique fail to completely solve the data imbalanced problem, which remains further research.

***5) Mis-classified Cases By MoT-LoGNN:*** The examples shown in Fig.7. are representative misclassified samples by the proposed MoT-LoGNN. The subtitle shows the actual label of the samples. It can be seen that these samples are extremely difficult to distinguish. Sample (a) is mis-classified is mainly due to that the fallen person is not completely in the monitoring area. As to sample (b), the fallen person with body curling is easily mis-classified as the standing person. Sample (c) shows a case where the MoT-LoGNN mis-classifies a standing person as a fallen person when the the background temperature is high enough. When two standing person are closed enough, the MoT-LoGNNmay wrongly judge them as a fallen person, as shown in sample (d).



(a) fallen (b) fallen

(c) non-fallen (d) non-fallen

Fig. 7. Mis-calssified samples of the MoT-LoGNN.

## V. Conclusions And Future Works

In this study, we use have proposed a robust approach that distinguishes falls and non-fall shapes from low-resolution images captured by low-cost and non-invasive thermal vision sensors. The device provides a zenithal point of view from the ceiling where it is located to detect falls in instances of both single and multi-occupancy. We propose the use of Convolutional Neural Networks to extract features from the thermal images and then use a Radial Basis Function Neural Network trained via the minimization of the Localized Generalization Error Bound as the classifier (T-LoGNN) to reduce the effects of thermal images with strong amounts of noise and blurred areas on the classification results. In addition, we propose a multi-occupancy fall detection method MoT-LoGNN in response to the decrease of classification accuracy caused by the increasing complexity of thermal images in the multi-occupancy scenarios. Experimental results demonstrated that the MoT-LoGNN achieved the best performance on both single and multi-occupancy scenarios.

However, the proposed MoT-LoGNN still have some limitations. The binarization threshold is set for a given data set in this work. Therefore, its fall detection performance may decline if the MoT-LoGNN is directly applied to other environments. Meanwhile, the proposed MoT-LoGNN conducts fall detection only using a single thermal image, where the decision making might be wrong owing to noise and blur in a thermal image. Thus, decision made using a sequence of successive thermal images rather than a single one should be more preferrable.

In the future, we will verify the proposed MoT-LoGNN in more real application scenarios. Research related to the image decomposition algorithm adapted to different environments is a meaningful research direction to optimize MoT-LoGNN. In addition, due to the uncertainty of thermal images acquired by this low-cost device, temporal information will be added to improve the robustness of the recognition system.

## References


[1] World Health Organization. "World report on ageing and health," *World Health Organization*, 2015.

[2] Z. He, *et al*, "Prevalence of multiple chronic conditions amongolder adults in florida and the united states: comparative analysis of theoneflorida data trust and national inpatient sample," Journal of medicalInternet research, vol. 20, no. 4, p. e137, 2018.

[3] D. Marikyan, *et al*, "A systematic reviewof the smart home literature: A user perspective," Technological Fore-casting and Social Change, vol. 138, pp. 139–154, 2019

[4] M. L. Shuwandy, *et al.*, "Sensor-based mHealth authentication for real-time remote healthcare monitoring system: A multilayer systematic review," *Journal of medical systems*, vol. 43, no.2, pp. 33, 2019.

[5] E. A. Kramarow, *et al.*, "Deaths from Unintentional Injury Among Adults Aged 65 and Over, United States, 2000-2013," no.2015. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2015.

[6] P. Vallabh and R. Malekian, "Fall detection monitoring systems: A comprehensive review," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 6, pp. 1809–1833, 2018.

[7] S. S. Kendri, *et al.*, "Development and monitoring of a fall detection system through wearable sensor belt," *Development*, vol. 6, no. 12, 2019.

[8] L. Wang, *et al*, "Pre-impact fall detection based onmulti-source cnn ensemble," *IEEE Sensors Journal*, vol. 20, no. 10, pp.5442–5451, 2020.

[9] J.-S. Lee and H.-H. Tseng, "Development of an enhanced threshold-based fall detection system using smart phones with built-in accelerometers," *IEEE Sensors Journal*, vol. 19, no. 18, pp. 8293–8302, 2019.

[10] L. Montanini, *et al*, "Afootwear-based methodology for fall detection," *IEEE Sensors Journa*l,vol. 18, no. 3, pp. 1233–1242, 2018.

[11] T. Xu, *et al*., "New advances and challenges of falldetection systems: A survey," *Applied Sciences*, vol. 8, no. 3, p. 418, 2018.

[12] S. M. Adnan, *et al*, "Fall detection through acoustic local ternary patterns," *Applied Acoustics*, vol. 140, pp. 296–300, 2018

[13] E. Cippitelli, *et al.*, "Radar andrgb-depth sensors for fall detection: A review," *IEEE Sensors Journal*, vol. 17, no. 12, pp. 3585–3604, 2017.

[14] W. K. Wong, *et al*., "Home alone faint detection surveillance system using thermal camera," in Computer Research and Development, 2010 Second International Conference on. IEEE, 2010, pp. 747–751.

[15] K.-S. Song, *et al*., "Histogram based fall prediction of patients using a thermal imagery camera," in 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI). IEEE, 2017, pp. 161-164.

[16] S. Moulik and S. Majumdar, "Fallsense: An automatic fall detectionand alarm generation system in iot-enabled environment," IEEE Sensors Journal, vol. 19, no. 19, pp. 8452–8459, 2018.

[17] P. Mazurek, *et al*, "Use of kinematic and mel-cepstrum-related features for fall detection based on data from infrareddepth sensors," Biomedical Signal Processing and Control, vol. 40, pp.102–110, 2018.

[18] J. Rafferty, *et al.*, "Fall detection through thermal vision sensing," in Ubiquitous Computing and Ambient Intelligence. Springer, 2016, pp. 84–90.

[19] A. Hayashida, *et al.*, "The use of thermal ir array sensor for indoor fall detection," in Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on. IEEE, 2017, pp. 594– 599.

[20] J. Medina-Quero, *et al.*, "Detection of falls from non-invasive thermal vision sensors using convolutional neural networks," in Multidisciplinary Digital Publishing Institute Proceedings, vol. 2, no. 19, 2018, p. 1236.

[21] A. Akula, *et al*., "Deep learning approach for human action recognition in infrared images," *Cognitive Systems Research*, vol. 50, pp. 146–154, 2018.

[22] J. Adolf, *et. al*, "Deep neural network based body posture recognitions and fall detection from low resolution infrared array sensor," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018, pp. 2394–2399.

[23] J. Nogas, *et al.*, "Fall detection from thermal camera using convolutional lstm autoencoder," in Proceedings of the 2nd workshop on Aging, Rehabilitation and Independent Assisted Living, IJCAI Workshop, 2018.

[24] J. Nogas, *et al*., "Deepfall: Non-invasive fall detection with deep spatio-temporal convolutional autoencoders," *Journal of Healthcare Informatics Research*, vol. 4, no. 1, pp. 50–70, 2020.

[25] R. Gade and T. B. Moeslund, "Thermal cameras and applications: asurvey,*"Machine vision and applications*, vol. 25, no. 1, pp. 245–262, 2014.

[26] M. E. Jaspers, *et al*., "The flir one thermal imager for the assessment of burn wounds:Reliability and validity study," *Burns*, vol. 43, no. 7, pp. 1516–1523,2017.

[27] S. Mashiyama, J. Hong, and T. Ohtsuki, "Activity recognition using low resolution infrared array sensor," in 2015 IEEE International Conference on Communications (ICC). IEEE, 2015, pp. 495–500.

[28] M. A. L opez-Medina, *et al.,* "Evaluation of convolutional neural networks for the classification of falls from heterogeneous thermal vision sensors," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, p. 1550147720920485,2020.

[29] J. Rafferty, *et al.,* "Sensor central: A research oriented, device agnostic, sensor data platform," in Ubiquitous Computing and Ambient Intelligence, S. F.Ochoa, P. Singh, and J. Bravo, Eds.Cham:Springer International Publishing, 2017, pp. 97–108.

[30] J. Rafferty, *et al.*, "A scalable, research oriented, generic, sensor data platform," *IEEE Access*, vol. 6, pp. 45 473–45 484, 2018.

[31] S. Suzuki, and Satoshi, "Topological structural analysis of digitized binary images by border following," Computer vision, graphics, and image processing, vol. 30, no. 1, pp. 32–46, 1985.

[32] Z. Zhou, *et al.*, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4761– 4772.

[33] J. Park, and I. W. Sandberg, "Universal approximation using radial-basis function networks," *Neural Computation*, vol. 3, no. 2, pp. 246-257, 2014

[34] D. S. Yeung, *et al.*, "Localized generalization error model and its application to architecture selection for radial basis function neural network," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1294–1305, 2007.

[35] D. S. Yeung, *et al.*, "Mlpnn training via a multiobjective optimization of training error and stochastic sensitivity," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 5, pp. 978–992, 2016.

[36] W.W.Y. Ng, *et al.*, "Diversified Sensitivity-Based Undersampling for Imbalance Classification Problems," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2402 – 4212, 2015.

[37] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5353–5360.